



Published in final edited form as:

Circulation. 2019 November 12; 140(20): 1610–1612. doi:10.1161/CIRCULATIONAHA.119.039912.

Emerging Affinity Reagents for High Throughput Proteomics: Trust, but Verify

Mark D. Benson, MD, PhD¹, Debby Ngo, MD¹, Peter Ganz, MD², Robert E. Gerszten, MD¹

¹Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA

²Division of Cardiology, Zuckerberg San Francisco General Hospital, University of California, San Francisco

High throughput proteomic profiling of human plasma is immensely challenging due to the extraordinary complexity and broad dynamic range of plasma protein concentrations. Mass spectrometry (MS) is unmatched for its ability to provide unambiguous identification of specific peptides, though its application to plasma proteomics at the population level remains prohibitive due to slow throughput and inability to quantify low abundant proteins without extensive upfront sample separation. Enzyme-linked immunosorbent assay (ELISA)-based approaches in turn have been limited by the unavailability of well-credentialed antibodies for most blood proteins and by issues with cross-reactivity, which significantly hampers the ability to multiplex antibodies for higher throughput analyses.

To address these challenges, emerging proteomics technologies have integrated new types of affinity reagents to bind proteins. Single-stranded DNA aptamers are one such technology that has been recently leveraged for large scale, human plasma profiling^{1, 2}. These short nucleotide segments (“aptamers”) fold into complex structures that bind protein targets through shape and charge complementarity with high affinity and specificity. Vast libraries that contain every possible sequence of these nucleotide segments can be randomly synthesized (typically containing $\sim 10^{14}$ - 10^{15} unique molecules) and exposed to a recombinant protein of interest. Through an iterative selection process, the aptamer with the best binding/capture properties can be selected and then amplified and tested for cross-reactivity, stability, and target binding properties.

In the assay, aptamers can be highly multiplexed to serve as “pull down” reagents for several thousand protein targets at a time from complex samples (blood, cerebrospinal fluid, tissue homogenate, etc.). Following a series of wash steps, each aptamer can then be released from its bound protein target, allowed to hybridize to a complementary sequence on a microarray DNA chip, and quantified by fluorescence. Thus, this technology efficiently translates protein concentrations to DNA concentrations, which can be measured using widely

Correspondence: Robert E. Gerszten, MD, Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02115. Phone(617) 667-8800. rgerszte@bidmc.harvard.edu.

Conflict of Interest Disclosures

MDB: none. DN: none. PG: serves on a medical advisory board to SomaLogic, Inc., for which he accepts no salary, honoraria, or any other financial incentives. REG: none.

available molecular genetics approaches. Currently, DNA aptamers can quantify the levels of about 5,000 different target proteins.

With this technology now viable, the current crucial challenge in the field is to develop approaches to validate the specificity and selectivity of each of these new reagents. Complementary approaches – leveraging “orthogonal” techniques including mass spectrometry, functional genomics, and alternative affinity-based approaches—can all be used to systematically validate currently available aptamer reagents.

In one such MS-based approach, for example, individual aptamers can be immobilized on beads, incubated with a study sample, and then selectively eluted for analysis by liquid chromatography (LC)-MS/MS. This not only allows for the determination of the limit of detection and limit of quantification of each aptamer of interest, but also provides important information in regard to the analytic specificity of the aptamer.

The application of high-throughput protein affinity reagents to thousands of human samples also affords new ways to leverage available genetic information to explore aptamer specificity. Genetic association studies can be conducted on circulating levels of each aptamer-measured protein. Detection of a genetic variant located within the gene that encodes the measured protein provides strong, “orthogonal” evidence that the aptamer is capturing its intended protein target, resulting in a valid quantitative assay. For example, we and others have demonstrated that a single nucleotide polymorphism (“SNP”, rs41271951) located within the cathepsin S gene is significantly associated with plasma levels of aptamer-measured cathepsin S protein—a cysteine protease involved in atherogenesis^{3, 4}. This provides strong supporting evidence that the cathepsin S aptamer is indeed binding its intended cathepsin S target protein. Further, these protein-SNP data can provide useful biological insight. For example, in addition to the rs41271951 SNP that is predicted to result in a missense substitution within the well-studied signal peptide responsible for cathepsin S secretion efficiency, additional associated SNPs point toward novel regulatory mechanisms of circulating cathepsin S levels in plasma.

Emerging technologies are also beginning to allow for multiplexed antibody-based assays to be used to validate aptamer reagents in a systematic fashion. For example, oligonucleotide-labelled antibodies have recently been incorporated into a novel, commercially available platform (Olink) that can be used to quantify levels of over a thousand human proteins. As there is approximately a 50% overlap between protein targets that are included on this platform and the SOMAscan DNA aptamer-based platform, investigators can now systematically compare these complementary approaches.

It is important to highlight that there are significant ongoing efforts to study and improve the specificity properties of novel affinity reagents, themselves. Specific chemical side chains can be added to DNA aptamers, for example, in order to generate reagents with very slow off-rates that can withstand vigorous wash steps and thus reduce non-specific binding. The inclusion of polyanionic competitor reagents during these wash steps (aptamers are also polyanions) further reduces binding of non-specific proteins to these reagents. Oligonucleotide-labelled antibody reagents require pairwise binding of two probes in order

to improve specificity. Leveraging these approaches, these affinity-based platforms can now discriminate between highly homologous proteoforms.

Despite these ongoing efforts, some investigators have recently suggested that emerging affinity reagents for high throughput proteomics are not valuable until specificity and additional performance characteristics match that of clinical assays. Such a demand on performance ignores the value-add of these reagents to high throughput *discovery*, which, if desired, can be confirmed with targeted approaches, as detailed above. The challenges facing new affinity proteomic tools are reminiscent of criticisms in the early GWAS era – when initial SNP arrays pointed only to loci of interest and clearly did not perform with CLIA-ready clinical performance standards. These obstacles were overcome through advances in probe design, quality control procedures, experimental and statistical methods, and orthogonal validation. Consequently, early skepticism of this technology has been replaced by recognition of its potential to inform disease biology.

These advances have allowed for significantly improved accessibility to proteomic profiling studies across a variety of experimental applications in large human cohorts. For example, investigators have recently applied the aptamer platform to nearly 2000 individuals across two large cohorts of patients with coronary heart disease to identify a signature of nine plasma proteins associated with subsequent CVD risk². Application of the same platform to nearly five hundred baseline plasma samples from participants of the ILLUMINATE trial recently identified a similar protein signal that predicted harm from the cholesterol ester transfer protein inhibitor torcetrapib⁵. While these clinical applications provide important proof of concept for real world clinical applications, validation in heterogeneous cohorts as well as assessment of how much this information adds to clinical findings are ongoing. Perhaps the most promising applications of these high throughput technologies relate to fundamental biology. By integrating proteomics and genomics, for example, investigators have begun to identify the genetic architecture of plasma proteins and to uncover “master regulators” of the circulating proteome and disease-associated effectors leveraging Mendelian randomization strategies^{3, 4}.

In sum, aptamers and oligonucleotide-labelled antibodies have opened the door to large-scale human proteomic profiling—with reproducibility and precision analogous to every molecular discovery tool. In parallel, continued appreciation for a multidisciplinary, collaborative approach, including public sharing of data, has contributed to remarkable progress to date with the next central challenge in the field focused on developing systematic methods to evaluate and improve the technical sensitivity and selectivity of individual reagents. With great promise yet clear recognition of affinity-based proteomics limitations and ongoing efforts to address them, we suggest a previously used guiding principle: “Trust, but verify...”

References

1. Ngo D; Sinha S; Shen D; Kuhn EW; Keyes MJ; Shi X; Benson MD; O’Sullivan JF; Keshishian H; Farrell LA; Fifer MA; Vasan RS; Sabatine MS; Larson MG; Carr SA; Wang TJ; Gerszten RE, Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. *Circulation* 2016, 134, 270–85. [PubMed: 27444932]

2. Ganz P; Heidecker B; Hveem K; Jonasson C; Kato S; Segal MR; Sterling DG; Williams SA, Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. *JAMA* 2016, 315, 2532–41. [PubMed: 27327800]
3. Benson MD; Yang Q; Ngo D; Zhu Y; Shen D; Farrell LA; Sinha S; Keyes MJ; Vasani RS; Larson MG; Smith JG; Wang TJ; Gerszten RE, Genetic Architecture of the Cardiovascular Risk Proteome. *Circulation* 2018, 137, 1158–1172. [PubMed: 29258991]
4. Sun BB; Maranville JC; Peters JE; Stacey D; Staley JR; Blackshaw J; Burgess S; Jiang T; Paige E; Surendran P; Oliver-Williams C; Kamat MA; Prins BP; Wilcox SK; Zimmerman ES; Chi A; Bansal N; Spain SL; Wood AM; Morrell NW; Bradley JR; Janjic N; Roberts DJ; Ouwehand WH; Todd JA; Soranzo N; Suhre K; Paul DS; Fox CS; Plenge RM; Danesh J; Runz H; Butterworth AS, Genomic atlas of the human plasma proteome. *Nature* 2018, 558, 73–79. [PubMed: 29875488]
5. Williams SA; Murthy AC; DeLisle RK; Hyde C; Malarstig A; Ostroff R; Weiss SJ; Segal MR; Ganz P, Improving Assessment of Drug Safety Through Proteomics: Early Detection and Mechanistic Characterization of the Unforeseen Harmful Effects of Torcetrapib. *Circulation* 2018, 137, 999–1010. [PubMed: 28974520]